

## Durham Research Online

---

### Deposited in DRO:

20 April 2016

### Version of attached file:

Accepted Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Joshi, S. and Srivastava, A. and Jermyn, I.H. (2007) 'Riemannian analysis of probability density functions with applications in vision.', in 2007 IEEE Conference on Computer Vision and Pattern Recognition ; proceedings. Piscataway, NJ: IEEE, pp. 1664-1671.

### Further information on publisher's website:

<http://dx.doi.org/10.1109/CVPR.2007.383188>

### Publisher's copyright statement:

© 2007 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

### Additional information:

## Use policy

---

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

# Riemannian Analysis of Probability Density Functions with Applications in Vision

Anuj Srivastava  
Department of Statistics  
Florida State University  
anuj@stat.fsu.edu

Ian Jermyn  
ARIANA Group, INRIA  
Sophia Antipolis, France  
ian.jermyn@sophia.inria.fr

Shantanu Joshi  
Department of Electrical Engineering  
Florida State University  
joshi@eng.fsu.edu

## Abstract

*Applications in computer vision involve statistically analyzing an important class of constrained, non-negative functions, including probability density functions (in texture analysis), dynamic time-warping functions (in activity analysis), and re-parametrization or non-rigid registration functions (in shape analysis of curves). For this one needs to impose a Riemannian structure on the spaces formed by these functions. We propose a “spherical” version of the Fisher-Rao metric that provides closed form expressions for geodesics and distances, and allows an efficient computation of statistics. We compare this metric with some previously used metrics and present an application in planar shape classification.*

## 1. Introduction

Several applications in computer vision, such as texture analysis, activity analysis and shape analysis, use mathematical representations involving a type of constrained, non-negative functions. To study observed variability within and across classes, one has to develop statistical models on appropriately constrained function spaces. Towards this goal, one needs to develop metrics, probability models, estimators and other tools for the desired inferences. The main difficulty here comes in performing calculus while respecting the nonlinear constraints imposed on these functions. A natural solution is to work on the nonlinear mani-

folds formed by the allowable functions and to utilize the underlying differential geometries of these manifolds to perform statistics. There are several mathematically equivalent choices of manifolds to pursue this idea. This general framework has previously been used by several researchers. However, questions remain about the choice of: (i) the representation and (ii) the Riemannian metric associated with that representation. In this paper we focus on a specific set of closely-related constrained functions – probability density functions (pdfs), warping functions, non-rigid registration functions, 1D diffeomorphisms, and re-parametrization functions – and study the different choices of representations. Our goal is to choose a space and a metric that allows an efficient computations of statistical tools for applications in computer vision.

Before we present different representations and metrics, we specify the functions of interest and their motivating applications. Perhaps the most important example is the use of pdfs in modeling frequencies of pixel values in images. Very commonly images are filtered using pertinent filters and the resulting images are used to estimate probability densities of the filtered pixel values. Applications of this tool include image retrieval [17], image compression [16], and texture synthesis [19, 14]. The second relevant problem is in *activity analysis* where one studies a time-series of individual events in order to classify this activity as a whole. For instance, consider the problem of a person arriving at the airport, checking in at an airline counter, and going to the departure gate to catch a flight. Since these individual events can be performed with random time

delays, one has to introduce a time-warping function in order to register and match observations [18, 15, 3]. The time-warping functions are naturally constrained to be non-decreasing functions and can be viewed as cumulative distribution functions (cdfs). With an additional positivity constraint, these functions form the group of 1D diffeomorphisms, whose 2D counterparts has famously been used in development of deformable template theory [9]. The third problem is in analyzing the shapes of closed, planar curves that are available as ordered sets of points. In order to compare them in a manner that is invariant to their parameterizations, one forms a quotient space, called the shape space, that is defined as the space of closed curves modulo all possible re-parameterizations [8, 11]. A related problem is to perform non-rigid registration of points across curves [12, 13].

Although the three above-mentioned applications are quite different, the sets of constrained non-negative functions used there are closely related. As described later, these functions may be represented and studied in several different but equivalent representations (pdfs, cdfs, log-densities, or square-root densities), each of which maybe better suited to particular tasks. An important step in classifying observations using these functions is to compute distances between any two arbitrary functions. This task is accomplished by imposing Riemannian structures on appropriate manifolds formed by these functions. The most natural Riemannian metric in this context is the so called *Fisher-Rao* metric that has been used extensively in computer vision [7, 6, 12, 13]. Čencov [2] showed that this is the only metric that is invariant to re-parametrizations of pdfs involved. This metric has also played an important role in information geometry due to the pioneering efforts of Amari [1].

The remaining question is: What choice of representation of functions (and hence the Fisher-Rao metric) is most efficient for our applications? We will demonstrate that the square-root form, defined to be the square-root of a pdf, results in the desired manifold to be a unit sphere inside a larger Hilbert space with the usual  $\mathbb{L}^2$  metric. In view of the spherical nature of the underlying space, many of the desired quantities (geodesics, exponential maps, inverse of exponential maps) are available in analytic forms. This is in contrast to the past usage of the Fisher-Rao metric where metrics and geodesics had to be approximated using numerical methods. In this paper, we will demonstrate the computational advantages of using the square-root form, and its associated Fisher-Rao metric, in vision applications.

The rest of this paper is organized as follows. Sec-

tion 2 presents three of several applications in computer vision that motivate this framework and introduces different mathematical representations of the functions of interest. Section 3 summarizes the differential geometry of the chosen representation and Section 4 demonstrates the computation of sample means using geometric tools. The paper ends with a demonstration of the proposed representation in a problem of binary shape classification.

## 2. Motivations & Representations

In this section we present some motivating applications for studying the type of constrained, non-negative functions that we have focused on. Additionally, we present several choices for representing these functions and discuss the structures of the resulting Riemannian manifolds equipped with the Fisher-Rao metric expressed in these representations.

### 2.1. Motivating Problems

We start by presenting some applications that involve the functions of interest.

1. **Spectral PDFs of Images:** As the first example, we highlight the use of pdfs in spectral analysis of texture, natural, or man-made images. Shown in Figure 1 is an example: the top left panel shows an image  $I$  that is then filtered using Gabor filters [4] at different orientations. For each resulting image, one computes a pdf of the gray scale pixel values. The remaining three panels shows examples of such pdfs. In the spectral analysis, each image is represented by a collection of pdfs, generated for a pre-determined bank of Gabor and other filters. Two images are compared by comparing their respective pdfs under the same filters. “Image templates”, denoting the central tendency of images in a class, can be defined as “averages” of the corresponding pdfs. Rescaling the image pixels to take values in the range  $[0, 1]$ , one is interested in tools for computing distances and averages on the set of pdfs on  $[0, 1]$ .
2. **Time Warping in Activity Analysis:** Consider the problem of activity analysis and classification, where one is interested in studying a pre-determined sequence of events that are performed with random time-separations. To compare different instances of the same activity, we need to use time-warping functions that allow registration of such instances. For example, define the activity of *catching a flight at an airport*. We define six sub-events: **A** - enter airport, **B** - join check-in line,

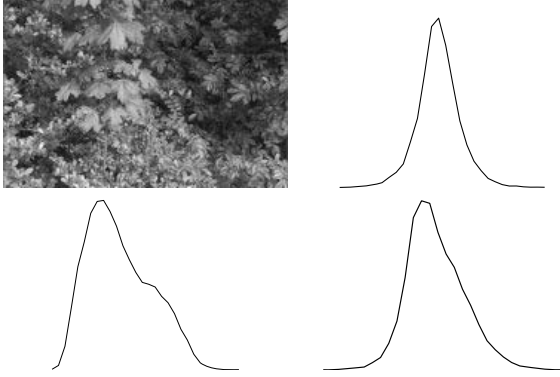


Figure 1. A natural image  $I$  (top left) and some pdfs of  $I * F(\theta)$ , where  $F$  is a Gabor filter with orientation  $\theta$ , for different  $\theta$ s.

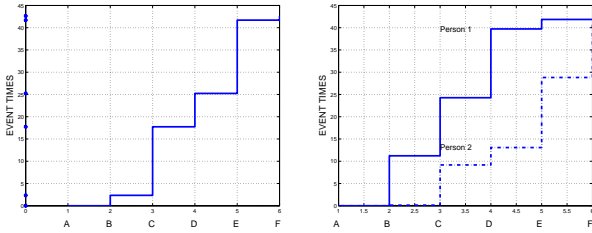


Figure 2. Left: An observation of an activity made of six events. Right: Two instances of this activity that have to be compared.

**C** - obtain boarding pass, **D** - clear X-ray check, **E** - enter terminal/gate area, and **F** - board airplane. An observation of this activity is denoted by  $T \equiv \{t_1, t_2, \dots, t_6\}$ , where  $t_i \in [0, 1]$  are the times of occurrences of the sub-activities; an example is shown in the left plot in Figure 2. For a time-warping function (made precise later)  $\phi$ , the set  $\phi(T) = \{\phi(t_i)\}$  becomes another occurrence of that activity with events occurring at  $\phi(t_i)$ . Since the sequence of events is maintained, one considers  $\phi(T)$  as the same activity as  $T$  and would like to cluster and classify it appropriately. To quantify differences between any two occurrences of the same activity,  $\phi_1(T)$  and  $\phi_2(T)$ , we will need to define and compute distances and statistics on the space of allowable  $\phi$ s.

### 3. Re-parameterizations in Shape Analysis:

Consider a simple, closed parameterized curve  $\alpha$  in a plane. For studying its shape, we resize  $\alpha$  to be of unit length. If  $\alpha$  is sampled differently, i.e. the sampled points are spaced differently, then the change in its representation is given by  $\alpha(\phi(s))$ , where  $\phi$  is a re-sampling function. Shown in Figure 3 is an example of three different parameteriza-

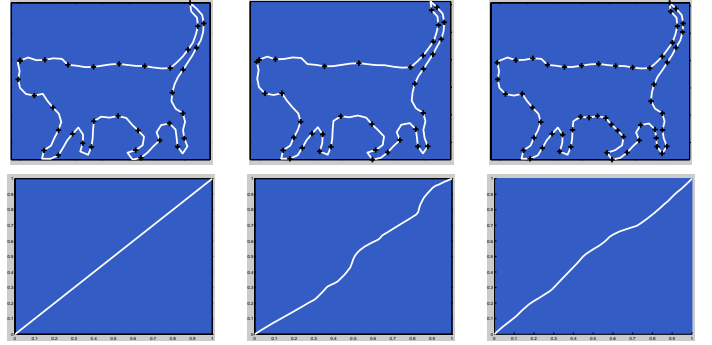


Figure 3. Same curve with three different parameterizations. The left one is sampled at uniform speed, or arc-length parametrization, while the other two have  $\phi$ s that are different from the identity.

tions of the same shape. The first one is the arc-length parametrization, i.e.  $\phi(s) = s$ , while the other are non-uniform speed parameterizations. Given two arbitrarily sampled closed curves  $\alpha_1$  and  $\alpha_2$ , we want to compare their shapes, taking all possible re-samplings into account. This problem has also been termed as non-rigid registration of points across shapes. In practical cases where shapes are observed in presence of noise, one needs to include probability models on space of  $\phi$ s to perform robust inferences.

## 2.2. Riemannian Representations

In the applications mentioned above, the functions of interest are, mathematically speaking, closely inter-related and can be represented in many ways. The ultimate choice of representation should be dependent on the ease of implementation in the ensuing application. A common issue in all these representations is that the underlying spaces are not vector spaces but are nonlinear (differentiable) manifolds; this creates a need to use tools from differential geometry – Riemannian metrics, geodesics, exponential maps, etc – on these manifolds for defining and computing statistics. As stated earlier, the choice of metric is fixed to be the Fisher-Rao metric because it is the only metric that is invariant to re-parameterization.

Next we enumerate different choices of representations and the associated forms of the Fisher-Rao metric:

1. **Probability density function  $p$ :** Each of the constrained, non-negative function of interest can be written as a pdf. To simplify discussion, we restrict to the space of pdfs on the interval  $[0, 1]$ ,

forming the set:

$$\mathcal{P} = \{p : [0, 1] \rightarrow \mathbb{R} | \forall s, p(s) \geq 0 \text{ and } \int_0^1 p(s) ds = 1\} .$$

$\mathcal{P}$  is not a vector space. The Fisher-Rao metric on  $\mathcal{P}$  can be stated as follows: for any point  $p \in \mathcal{P}$  and the tangent vectors  $v_1, v_2 \in T_p(\mathcal{P})$ , the inner-product is given by

$$\langle v_1, v_2 \rangle = \int_0^1 v_1(s) v_2(s) \frac{1}{p(s)} ds . \quad (1)$$

Here,  $T_p(\mathcal{P})$  is the set of functions tangent to  $\mathcal{P}$  at  $p$ . Amongst all possible representations, the pdf turns out to be one of the most difficult representations to work with. The main difficulty comes from the need for ensuring  $p(s) \geq 0$  for all  $s$ . For example, in case one is trying to compute a geodesic between any two elements of  $\mathcal{P}$ , it is quite difficult to ensure that  $p$  remains non-negative along the whole path. As an aside, we remark that the path  $tp_1 + (1-t)p_2$ , for  $0 \leq t \leq 1$  and  $p_1, p_2 \in \mathcal{P}$ , is not a geodesic between  $p_1$  and  $p_2$  under the Fisher-Rao metric.

2. **Cumulative distribution function  $\phi$ .** Associated with each element of  $\mathcal{P}$  is a unique cdf  $\phi(s) = \int_0^s p(t) dt$ .  $\phi$  is a differentiable mapping from  $[0, 1]$  to itself. Additionally, if  $p > 0$  then  $\phi$  is also an invertible map. Define the set of all cdfs:

$$\Phi = \{\phi : [0, 1] \rightarrow [0, 1] | \forall s, \dot{\phi}(s) > 0, \phi(0) = 0, \phi(1) = 1\} .$$

$\Phi$  forms a group with the group operation given by composition, i.e.  $\phi_1, \phi_2 \in \Phi$ , the group operation is given by  $\phi_2(\phi_1(s))$ . The identity element of  $\Phi$  is the function  $id(s) = s$ . The Fisher-Rao metric for this representation is given by: for any  $v_1, v_2 \in T_\phi(\Phi)$ , we have

$$\langle v_1, v_2 \rangle = \int_0^1 \dot{v}_1(s) \dot{v}_2(s) \frac{1}{\dot{\phi}(s)} ds . \quad (2)$$

$\Phi$  is somewhat easier than  $\mathcal{P}$  to analyze in view of its group structure. Also, note that the time-warping functions in activity analysis and the re-parametrization functions (or non-rigid registration functions) in shape analysis can be directly written as elements of  $\Phi$ . We remark that  $\Phi$  is also the space of 1D diffeomorphisms of the interval  $[0, 1]$ .

3. **Log density function:** Several past papers have used the logarithm of  $p$  to represent and analyze

probability densities. Of course, this representation requires the function  $p$  to be strictly positive. The corresponding representation space is:

$$\mathcal{L} = \{\nu : [0, 1] \rightarrow \mathbb{R} | \int_0^1 \exp(\nu(s)) ds = 1\} .$$

The Fisher-Rao metric on this representation is given by: for  $v_1, v_2 \in T_\nu(\mathcal{L})$ :

$$\langle v_1, v_2 \rangle = \int_0^1 v_1(s) v_2(s) \exp(\nu(s)) ds . \quad (3)$$

Although this representation has shown some success in texture analysis [10], there are a few major technical limitations here. Firstly, the pdf  $p$  should be strictly positive to have a logarithmic representation. Secondly, the Riemannian structure on the space is complicated and one has to use numerical techniques to compute geodesics on this space. For example, Mio et al. [10] use a shooting method to find geodesics between any two log-density functions. As demonstrated through an example later, this approach often results in paths that may not reach the target function and thus leads to large errors.

4. **Square-root density function:** The final choice of representation is the square root function:  $\psi = \sqrt{p}$ . Due to the nature of the square root, this function is not a unique representation of  $p$ ; uniqueness can be imposed by assuming  $\psi$  to be a non-negative function. Note that there is no requirement for  $p > 0$  for this representation to work. Here one considers the space:

$$\Psi = \{\psi : [0, 1] \rightarrow \mathbb{R} | \psi \geq 0, \int_0^1 \psi^2(s) ds = 1\} .$$

For any two tangent vectors  $v_1, v_2 \in T_\psi(\Psi)$ , the Fisher-Rao metric is given by:

$$\langle v_1, v_2 \rangle = \int_0^1 v_1(s) v_2(s) ds . \quad (4)$$

Eqn. 4 shows that the space of square-root densities can be viewed as the non-negative orthant of the unit sphere in a Hilbert space. In this larger space, not only are the inner products of tangent vectors defined, but also the inner products of the elements of the Hilbert space itself (and hence of elements of  $\Psi$ ). The distance in the larger space between two elements is simply the norm of their difference. Note that this is not the same distance in the space  $\Psi$ , which is the unit sphere.

It must be noted that the metrics in Eqns. 1-4 are all actually equivalent after appropriate changes of variables. They have simply been expressed in different coordinate systems. Additionally, they are invariant to re-parametrization.

**Theorem 1** *The Fisher-Rao metric is invariant to re-parameterizations.*

**Proof:** We prove this using the square-root representations but the proof is similar for all other representations. Let  $v_1, v_2 \in T_\psi(\Psi)$  for some  $\psi \in \Psi$  and  $\phi \in \Phi$  be a re-parametrization function. The re-parametrization action takes  $\psi$  to  $\psi(\phi)\sqrt{\dot{\phi}}$  and  $v_i$  to  $\tilde{v}_i \equiv v_i(\phi)\sqrt{\dot{\phi}}$ . The inner product after re-parametrization is given by:

$$\begin{aligned} \int_0^1 \tilde{v}_1(s)\tilde{v}_2(s)ds &= \int_0^1 v_1(\phi(s))\sqrt{\dot{\phi}(s)}v_2(\phi(s))\sqrt{\dot{\phi}(s)}ds \\ &= \int_0^1 v_1(t)v_2(t)dt, \quad t = \phi(s). \end{aligned}$$

which is the same as before re-parametrization and, hence, invariant.

Which of these different representations should one choose for texture, activity, and shape analysis? In this paper, we propose the use of square-root form for Riemannian analysis of constrained functions. The biggest advantage of this choice is that the resulting space  $\Psi$  is simply a unit sphere inside a larger Hilbert space with the  $\mathbb{L}^2$  metric. The differential geometry of a sphere is well understood. There are closed form expressions for computing geodesics, exponential maps, inverse exponential and, consequently, sample statistics on a unit sphere. The condition that  $\psi \geq 0$  is not too demanding; this amounts to restricting to a positive orthant of a unit sphere and does not impose any additional computational burden.

### 3. Differential Geometry of $\Psi$

In this section we specify the formulae for computing geodesics and other geometric quantities needed in vision applications. As stated above, the main advantage of selecting  $\Psi$  for analysis is that it is a convex subset of a unit sphere in  $\mathbb{L}^2$  and many of the geometric expressions are already known.

- **Geodesic Distance:** Given any two functions  $\psi_1$  and  $\psi_2$  in  $\Psi$ , the length of the geodesic connecting them in  $\Psi$  is given by:

$$d(\psi_1, \psi_2) = \cos^{-1}\langle\psi_1, \psi_2\rangle \quad (5)$$

where the inner product is as defined in Eqn. 4, but now applied to the elements of  $\Psi$  rather than the tangent vectors.

- **Geodesic:** The geodesic between two points is easily derived by noting that the radial projection to unit norm of the straight line joining the two points is the geodesic between the two points. The straight line is given by  $\tilde{\psi}(t) = (1-t)\psi_0 + t\psi_1$ . The projection to  $\Psi$  is then

$$\psi(t) = \frac{(1-t)\psi_1 + t\psi_2}{t^2 + (1-t)^2 + 2t(1-t)\langle\psi_1, \psi_2\rangle},$$

where it should be noted that  $t$  is not the arc-length parameterization.

Conversion to the distance parameterization leads to the following expression for the geodesic:

$$\psi(t) = \frac{1}{\sin(s_{12})} [\sin(s_{12} - t)\psi_1 + \sin(t)\psi_2], \quad (6)$$

where  $\cos(s_{12}) = \langle\psi_1, \psi_2\rangle$  is the cosine of the geodesic distance between the two points.

- **Exponential Map:** The geodesic can also be parameterized in terms of a direction in  $T_{\psi_1}(\Psi)$ :

$$G_t(v) = \cos(t)\psi_1 + \sin(t)\frac{v}{|v|}, \quad (7)$$

where  $v \in T_{\psi_1}(\Psi)$  and  $\langle v, \psi_1 \rangle = 0$ . As a result, the exponential map,  $\varepsilon : T_{\psi_1}(\Psi) \rightarrow \Psi$ , has a very simple expression:

$$\exp_{\psi_1}(v) = \cos(|v|)\psi_1 + \sin(|v|)\frac{v}{|v|}. \quad (8)$$

The exponential map is a bijection if we restrict  $|v|$  so that  $|v| \in [0, \pi)$ .

- **Inverse Exponential Map:** For any  $\psi_1, \psi_2 \in \Psi$ , we define  $v \in T_{\psi_1}(\Psi)$  to be the inverse exponential of  $\psi_2$  if  $\exp_{\psi_1}(v) = \psi_2$ ; we will use the notation  $\exp_{\psi_1}^{-1}(\psi_2) = v$ . This is computed using the following steps:

$$\begin{aligned} u &= \psi_2 - \langle\psi_2, \psi_1\rangle\psi_1 \\ v &= u \cos^{-1}(\langle\psi_1, \psi_2\rangle) / \sqrt{\langle u, u \rangle}. \end{aligned} \quad (9)$$

Since we have simple analytical expressions for computing these quantities, the resulting statistical analysis of elements of  $\Psi$  is much simpler compared to the other representations. For instance, the geometry of  $\mathcal{L}$  (using log-density coordinates) is too complicated to derive analytical expression for geodesics. Instead, one uses a numerical approach. Mio et al. [10] use a *shooting approach* for constructing geodesics on  $\mathcal{L}$ . The main idea is, given two log-densities  $\nu_1$  and  $\nu_2$  in  $\mathcal{L}$ , to find a tangent direction  $v \in T_{\nu_1}(\mathcal{L})$  such that a geodesic



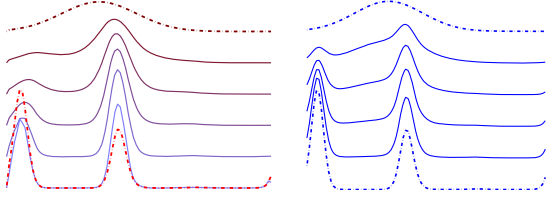


Figure 4. Limitations of shooting method. Computation of a geodesic using the shooting may not reach the target pdf (left picture) while no such problem exists for the analytical form (right picture) available for the square-root density representation. Geodesics are displayed using the corresponding elements in  $\mathcal{P}$  for convenience.

along  $v$  (constructed numerically) reaches  $\nu_2$  in unit time. This optimal direction  $v$  is found by minimizing a *miss error*, defined as the Euclidean distance between the function reached (for the current  $v$ ) and  $\nu_2$ . There are several disadvantages associated with this numerical approach. Firstly, one may not always be able to solve this minimization problem globally using numerical techniques. Secondly, the resulting geodesic may get close to  $\nu_2$  but not quite reach it.

Figure 4 highlights the problem in using the shooting method in forming a geodesic under the log-density representation. The left picture shows a path that has been computed using the shooting method in the space  $\mathcal{L}$  of the log-density functions, while the right path is computed using analytical expressions in the space  $\Psi$  of the square-root functions. (All the functions are displayed in terms of their pdfs for comparisons.) In each panel, the top and the bottom curves are the given densities  $p_1$  and  $p_2$ , while the intermediate curves denote equally spaced points on the geodesics. In the left picture, the geodesic starting from  $\nu_1 = \log(p_1)$  never quite reaches the point  $\nu_2 = \log(p_2)$ . In comparison, the geodesic computed using the analytical expression available for the square-root coordinates (right picture) has no such problem. Figure 5 shows some additional examples of geodesic paths between elements of  $\Psi$  (displayed using the corresponding elements in  $\mathcal{P}$ ).

#### 4. Sample Statistics on $\Psi$

An important ingredient in the statistical analysis of constrained functions is the computation of sample statistics such as means and covariances. For example, given a few observations of an activity, each resulting from a different time-warping function, one is interested in computing a *template* of that activity that involves taking average of all the observed time-warping functions. Similarly, given a collection of pdfs from a set of images (under the same filter), we may use an

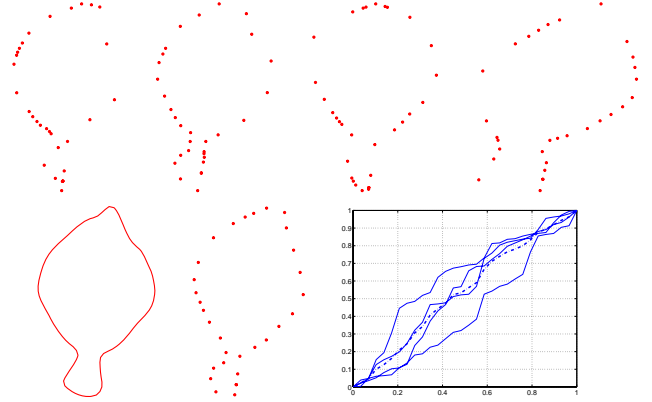


Figure 7. Top row: Four observations of a shape (shown in bottom left) at randomly sampled points, i.e. with arbitrary speed functions  $\phi$ s shown in bottom right. Bottom row: the middle panel shows this shape sampled at the mean  $\phi$ ; the mean  $\phi$  is shown on the right in broken line.

“average” pdf to characterize the typical statistics of this set. To define and compute means, we use the notion of Karcher mean [5] as follows: For a number of observations  $\psi_1, \psi_2, \dots, \psi_n$ , define their Karcher mean as:  $\bar{\psi} = \operatorname{argmin}_{\psi \in \Psi} \sum_{i=1}^n d(\psi, \psi_i)^2$ , where  $d$  is taken to be the geodesic distance on  $\Psi$ . The search for  $\bar{\psi}$  is performed using a gradient approach where an estimate is iteratively updated according to:

$$\mu \rightarrow \exp_{\mu}(\epsilon v), \quad v = \frac{1}{n} \sum_{i=1}^n \exp_{\mu}^{-1}(\psi_i)$$

where  $\exp$  and  $\exp^{-1}$  are given in Eqns. 8 and 9, respectively. The scalar  $\epsilon > 0$  is a step size for iteration and is generally taken to be smaller than  $\frac{1}{2}$ .

Shown in Figure 6 are some examples of  $\psi$ s and their Karcher means. As earlier, all the displays are in the form of the pdfs  $p = \psi^2$ . In this particular example, the original pdfs are made of Gaussians with different means and variances, although that parametric form has not been utilized in computing the Karcher means.

Another example for computing averages, in context of shape analysis, is shown in Figure 7. The top row shows four randomly sampled versions of the shape shown in the bottom left. The bottom middle panel shows this shape sampled using a  $\phi$  that is average of the four individual  $\phi_i$ s. These sampling functions are shown in the bottom right panel.

#### 5. Experiment on Shape Classification

To demonstrate the strength of this framework, we consider a simple shape classification problem. We restrict to the binary case where we observe one of the

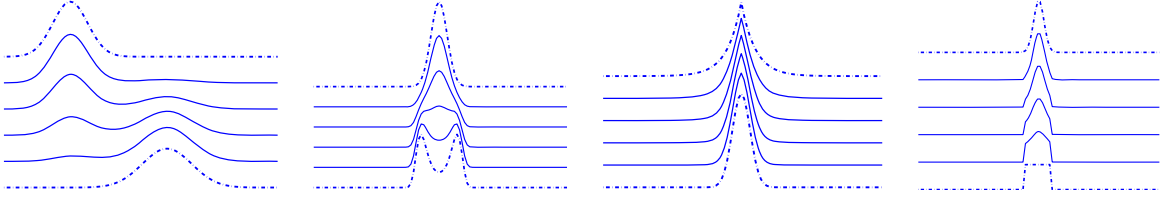


Figure 5. Geodesic paths in  $\Psi$  between some interesting square-root densities. All functions are displayed using their corresponding values in  $\mathcal{P}$  for convenience.

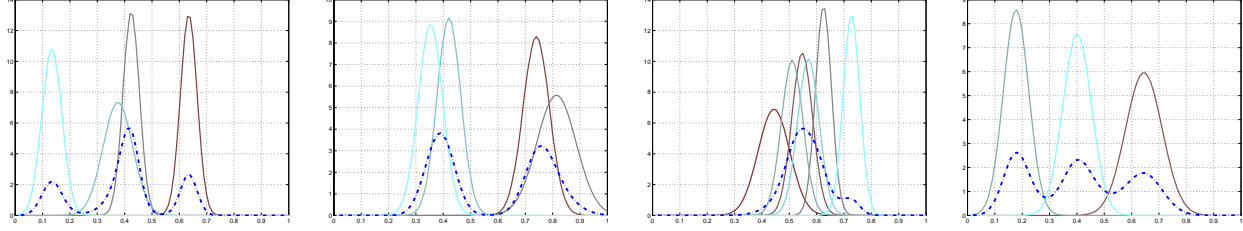


Figure 6. Examples of Karcher means of elements of  $\Psi$ . Each panel shows some Gaussian densities  $p_i = \psi_i^2$  with different means and variances in solid lines. Superimposed on them in broken line is their Karcher mean  $\bar{p} = \bar{\psi}^2$ .

two shapes at random sampling and in presence of additive Gaussian noise. The observed data is given by:

$$\alpha^d = \alpha_t(\gamma_0) + n, \quad t = 1, 2, \quad (10)$$

where  $\alpha_t$  is the true underlying shape and  $n$  is taken to be white Gaussian process, i.e.  $n_1(s), n_2(s)$  are independent Gaussian random variables and  $\gamma_0 \in \Phi$  denotes a random sampling involved in observing the template  $\alpha_t$ . Shown in Figure 8 is a pictorial example of this setup. The top row shows the two templates,  $\alpha_1$  and  $\alpha_2$ , associated with the two classes and the middle row shows four examples of  $\alpha^d$  with increasing observation noise from left to right.

For a given observation, the posterior probability that it belongs to the  $i^{th}$  class is computed as follows.

$$P(i|\alpha^d) \propto P(i) \int_{\gamma \in \Phi} P(\alpha^d|\gamma, i) P(\gamma|i) d\gamma, \quad (11)$$

The likelihood function can be taken to be  $P(\alpha^d|\gamma, i) \propto \exp(-E_i)$ , where

$$E_i = \|\alpha^d - \alpha_i(\gamma)\|^2. \quad (12)$$

The choice of  $P(\gamma|i)$  is more interesting. Using techniques presented in Section 4, we can compute the average re-parametrization function  $\bar{\gamma}_i$  associated with prior observations in each class and form a prior  $P(\gamma|i) \propto e^{-d(\gamma, \bar{\gamma}_i)^2}$ . Another possibility is to use the term  $e^{-d(\gamma, id)^2}$  which uses the squared distance from the identity in  $\Phi$  as the prior energy. This term is computed using the square-root representations of  $\gamma$

and  $id$  in  $\Psi$  and then using the geodesic distance between them in  $\Psi$  (Eqn. 5).

The integral in Eqn. 11 is often approximated by evaluating the integrand at the maximum likelihood estimate of  $\gamma$ . Let  $\hat{\gamma}_i$  be the optimal re-parametrization of the template  $(\phi_i, \theta_i)$  according to:

$$\hat{\gamma}_i = \underset{\gamma \in \Phi}{\operatorname{argmin}} \|\alpha^d - \alpha_i(\gamma)\|^2.$$

Then, the posterior probability is approximated by the term:

$$P(i|\alpha^d) \approx P(i) P(\alpha^d|\hat{\gamma}_i, i) P(\hat{\gamma}_i|i).$$

The bottom row of Figure 8 shows the results of this binary classification performance in two cases: one when only the likelihood term in Eqn. 12 is used (broken line) and when the full posterior is used (solid line). The left plot shows the case when  $\gamma_0$  is much closer to  $id$  than the case shown on the right. The classification performance was computed using 1000 Monte Carlo trials at each noise level. This experiment clearly shows the utility of the prior term  $P(\gamma|i)$  in Eqn. 12 in the classification process. Furthermore, our framework computes this distance very efficiently as a distance on a unit sphere.

## 6. Summary

We have proposed a “spherical” version of the Fisher-Rao metric for imposing Riemannian structure on a collection of related spaces: the space of pdfs, time-warping functions, re-parametrization functions, etc. The proposed metric is both computationally and



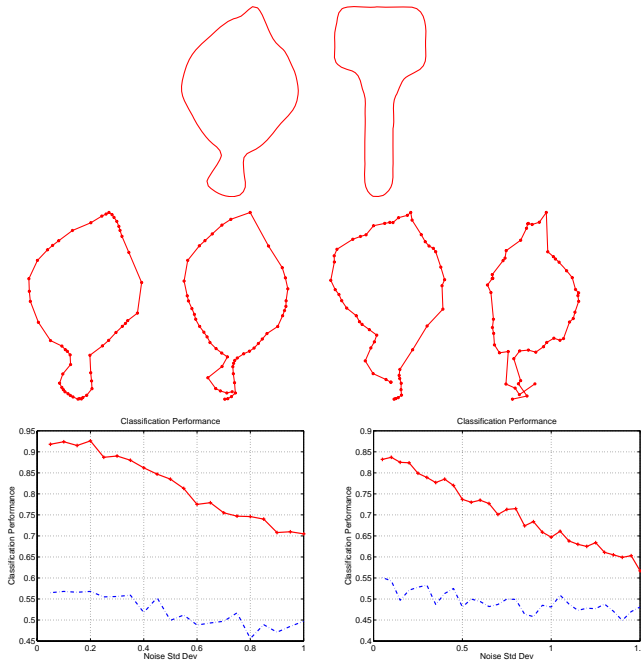


Figure 8. Top row: Templates for the two object classes. Middle row: Observations of class 1 template at random  $\gamma_0$ s and increasing noise from left to right. Bottom row: Plot of classification performance versus the observation noise. Left panel shows the case when  $\gamma_0$  is much closer to  $id$  and the right panel shows the opposite case. The broken line denotes the maximum-likelihood solution and the solid line denotes the maximum a-posteriori solution.

analytically much simpler, and allows efficient computation of statistics on a larger class of functions than previously used metrics. We have compared this metric with some previously used metrics and have presented an application in planar shape classification.

## References

- [1] S.-I. Amari and H. Nagaoka. *Methods of Information Geometry*. Oxford University Press, 2000. 2
- [2] N. N. Čencov. *Statistical Decision Rules and Optimal Inferences*, volume 53 of *Translations of Mathematical Monographs*. AMS, Providence, USA, 1982. 2
- [3] T. Darrell, I. Essa, and A. Pentland. Task-specific gesture analysis in real-time using interpolated views. *IEEE Trans. on PAMI*, 18(12):1236–1242, 1996. 2
- [4] B. Jahne, H. Haubecker, and P. Geibler. *Handbook of Computer Vision and Applications, Volume 2*. Academic Press, 1999. 2
- [5] H. Karcher. Riemann center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30:509–541, 1977. 6
- [6] S. J. Maybank. Detection of image structures using the Fisher information and the Rao metric. *IEEE Trans-*

- actions on Pattern Analysis and Machine Intelligence*, 26(12):1579–1589, 2004. 2
- [7] S. J. Maybank. The Fisher-Rao metric for projective transformations of the line. *Int. J. Comput. Vision*, 63(3):191–206, 2005. 2
- [8] P. W. Michor and D. Mumford. Riemannian geometries on spaces of plane curves. *Journal of the European Mathematical Society*, 8:1–48, 2006. 2
- [9] M. I. Miller and L. Younes. Group actions, homeomorphisms, and matching: A general framework. *Intern. Journal of Computer Vision*, 41(1/2):61–84, 2002. 2
- [10] W. Mio, D. Badlyans, and X. Liu. In *EMMCVPR 2005*, volume 3757 of *Lecture Notes in Computer Science*, pages 18–33. Springer, 2005. 4, 5
- [11] W. Mio and A. Srivastava. Elastic string models for representation and analysis of planar shapes. In *Proc. of IEEE Computer Vision and Pattern Recognition, Washington D.C.*, June, 2004. 2
- [12] A. Peter and A. Rangarajan. A new closed-form information metric for shape analysis. In *Proc. of MICCAI, Copenhagen*, 2006. 2
- [13] A. Peter and A. Rangarajan. Shape matching using the fisher-rao riemannian metric: Unifying shape representation and deformation. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2006. 2
- [14] J. Portilla and E. P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int'l Journal of Computer Vision*, 40(1):49–71, 2000. 1
- [15] C. Rao, M. Shah, and T. F. Syeda-Mahmood. Invariance in motion analysis of videos. In *Proceedings of the Eleventh ACM International Conference on Multimedia*, pages 518–527, 2003. 2
- [16] J. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Trans Sig Proc*, 41(12):3445–3462, December 1993. 1
- [17] E. Spellman, B. C. Vemuri, and M. Rao. Using the KL-center for efficient and accurate retrieval of distributions arising from texture images. In *Proc. of IEEE CVPR*, volume 1, pages 111–116, 2005. 1
- [18] A. Veeraraghavan and A. K. Roy-Chowdhury. The function space of an activity. In *Proc. of IEEE CVPR*, volume 1, pages 959–968, 2006. 2
- [19] S. C. Zhu, Y. N. Wu, and D. Mumford. “Minimax entropy principles and its application to texture modeling”. *Neural Computation*, 9(8):1627–1660, November 1997. 1